基于异常值识别卡尔曼滤波器的短期交通流预测 *

白伟华 [1]，张传斌 [1]，张塽旖 [2]，周 腾 [2]

(1. 肇庆学院 计算机科学与软件学院、大数据学院，广东 肇庆 526061; 2. 汕头大学计算机系，广东 汕头 515063)

摘 要：针对智慧交通的需求，提出了一种新颖有效的短时交通流预测方法，通过异常值识别扩展了卡尔曼滤波，使其能对噪声进行识别和过滤——异常值识别卡尔曼滤波器。利用卡尔曼滤波能有效地过滤导致系统不确定性的交通流波动，但这可能会使指示交通流突变的细微线索丢失。为了提升预测精度，应用离散小波变换对原始信号进行识别处理，在去掉异常值的同时保留原有对预测有效的信号源信息，此外还使用了历史参考值对预测值进行修正。最后，在四个基准数据集上的大量实验表明，与常用及最新的预测模型相比，其结果 MAPE 平均降低了 2.919%，RMSE 平均降低了 79.582。

关键词：卡尔曼滤波；有源噪声控制；短时交通流预测；状态向量

中图分类号：TP393.09 doi: 10.19734/j.issn.1001-3695.2020.03.0070

Outlier-identified Kalman filter for short-term traffic flow forecasting

Bai Weihua[1], Zhang Chuanbin[1], Zhang Shuangyi[2], Zhou Teng[2]

(1. *School of Computer Science & Software*、*School of Big Data*, *ZhaoQing University*, *Zhaoqing Guangdong* 526061, *China*; 2. *Dept. of Computer Science*, *Shantou University*, *Shantou Guangdong* 515063, *China*)

Abstract: A novel and effective technique for short-term traffic flow forecasting is presented. The main contribution is an extension of Kalman filter, such that it becomes to be able to identify the outlier and then filter out it; the present technique is named as Outlier-identified Kalman filter. The fluctuations of the traffic flow, which leads the system uncertainty, is often filtered out by the classic Kalman filter, and the subtle clues indicating the sudden change of the traffic flow may lose in this operation. To achieve better forecasting accuracy, discrete wavelet transform is used to process the original signal and the useful signals are preserved while de-noising. In addition, historical reference values are applied to correct predicted values. Extensive experiments on four benchmark datasets demonstrate an average improvement of 2.919% in MAPE and 79.58 in RMSE.

Key words: Kalman filter; active noise control; short-term traffic flow forecasting; state vector

0 引言

结合物联网、车辆网和智慧交通的应用，准确及时的交通流信息对智能交通系统(intelligent transport system, ITS)的成功部署至关重要，它能为物流部门、商业机构、旅游服务机构和政府管理机构提供可靠的交通流信息：优化路线规划、提高效率且节约时间；缓解交通拥堵，降低交通运输成本；减少汽车尾气排放污染、节省能源；减少交通事故，并提高交通运营效率。

为提升交通运输体系的智能化水平，应用大数据、移动互联网等技术与交通运输深度融合，交通流量是判别道路交通状态的一个关键因素。短期交通流预测是微观意义上的，与中观和宏观意义上以小时、天、甚至是年计算的基于交通规划的战略预测有本质区别：其主要内容是基于当前和过去的交通数据，采用适当的方法去滚动预测未来几秒到半小时的交通状况。然而，由于交通流内在的随机性和外部噪声[1~3]的存在，比如意外事故或人工交通管制[4,5]，找到既稳健又准确的预测算法是一项具有挑战性的任务。

针对交通流预测研究，各国学者已提出了 300 多种短时交通流预测模型和方法，主要包括线性理论模型、非线性理论模型、动态分配模型及基于知识发现的模型四大类。卡尔曼滤波器(Kalman filter, KF)及其变体由于具有计算效率高，所需内存少，适用于静态、非平稳数据分析，可使用实时数据动态更新其状态变量，并能更好地适应交通流量的变化等特性，是交通流预测任务中广泛使用的强大工具之一[6~8]。为了解决交通流中存在的高度不确定性(气候、交通事故、突发事件等)、简单的噪声过滤预处理对指示交通流突变的失真问题，本文提出了一种能够准确区分有效信号和噪声的卡尔曼滤波器，命名为异常值识别卡尔曼滤波器(outlier-identified Kalman filter, OiKF)。本文将 OiKF 应用于短期交通流预测，并在四个基准数据集上对其进行了实验评估，然后与经典的卡尔曼滤波器以及其他常用的参数和非参数技术进行性能比较，经大量实验证明，该技术提高了预测的精确度和准确性，证明了异常值识别卡尔曼滤波器在短期交通流预测方面的优越性。

1 相关研究

传统的卡尔曼滤波及其变体为分析和解决一大类估计问题提供了清晰强大的基本工具，被广泛应用于许多工程领域，

收稿日期：2020-03-25；**修回日期：**2020-05-08 **基金项目：**国家自然科学基金资助项目(61902232)；广东省自然科学基金资助项目(2018A030313291)、广东省教育科学规划项目(2018GXJK048)；广东省普通高校特色创新项目(自然科学，2019KTSCX199)；广东大学生科技创新培育专项资金资助项目(pdjh2020b0222)；肇庆学院校级科研项目(2019012612)

作者简介：白伟华(1976-)，男，广东高要人，副教授，博士，主要研究方向为任务序列分析、服务计算、大规模任务调度、多服务器控制(bandwer@163.com)；张传斌(1990-)，男，广东肇庆人，实验员，硕士，主要研究方向为计算智能，大数据，系统建模与优化；张塽旖(1996-)，女，青海西宁人，硕士研究生，主要研究方向为城市计算、深度学习；周腾(1987-)，男，广东汕头人，讲师，博士，主要研究方向为集成学习、深度学习、智能交通系统和医学图像处理.

特别是对于离散时间系统状态估计，包括信号处理，最优控制，导航等。实际应用包括参数估计，系统识别，目标跟踪，同步定位等。这些问题的很多过程都可以用状态空间系统去描述，如果本文能对一个过程建立数学模型，本文就可以利用数学工具去控制这个过程，并获得关于该过程的信息。利用状态空间系统对短期交通流预测问题进行建模已有不少研究。Okutani 和 Stephanedes[9] 首先在城市交通量预测中引入状态空间模型来预测高速公路入口坡道区域的交通分流，并验证该模型优于二代城市交通控制系统(UTCS-2)中使用的交通量预测模型。Stathopoulos 和 Karlaftis[10]也提出了一种基于最优状态估计理论的模型，并表明多变量状态空间模型的使用提高了 ARIMA 模型的预测精度。Hu 等人[11]提出了混合粒子群优化支持向量机回归模型(Hybrid PSO-SVR)，使用支持向量机回归模型进行短期交通流预测，并通过粒子群算法优化向量机回归模型的参数，减少模型学习时间。Guo 等人[12]提出使用一阶灰色模型 GM 进行短期城市道路交通流预测，该模型引入道路的负载和时延因素，与原始灰色模型相比，预测精度明显提高。最小平方提升模型(least squares boosting, LSBOOST)是一个集成多个学习算法的框架[13]，在进行短期电力需求预测时，具有良好效果，可尝试用于其他领域的短期预测。然而，如前文所述，短期交通流数据经常被局部噪声破坏，这可能会显著影响短期交通流的预测准确性。在先前的研究中，只关注了卡尔曼滤波在短期交通流预测中过滤噪声的作用，忽视了过滤噪声有损信号质量，例如，Xie 等人[14]研究了卡尔曼滤波器在离散小波分析中的应用，得出噪声污染会严重影响交通流预测性能的结论。Wang 等人[15]使用了一种基于卡尔曼滤波器的自适应交通流预测算法，该算法同时考虑了状态和观测高斯噪声。为了进一步提高预测精确度，Zhou 等人[7]通过补偿初步预测的误差，提出了一种用于交通流预测的新型混合卡尔曼滤波器。然而随着过滤噪声的能力逐步地提升，预测精确度趋于平稳，甚至开始下降。主要原因是，在过滤噪声的同时，有用的信号，尤其是部分指示交通流突变的细微线索也被过滤掉了。

与已有工作不同的是，本文利用一个新的成本函数重新构造了卡尔曼滤波器，提出了一种新颖有效的噪声识别卡尔曼滤波器 OiKF，该技术能够在实现高效噪声过滤的同时获得高质量有效信号，并将其应用于短期交通流预测，提出了一种鲁棒交通流预测算法。具体地，本文通过离散小波变换(discrete wavelet transform, DWT)将原始交通流数据分解重构成近似部分和细节部分。与只抛弃细节部分的 Zhou 等人[7]的工作相比，该方法能够考虑多级细节部分；本文将不同层次的细节部分作为影响预测性能的噪声或提高预测性能的线索，与保持了原始交通流数据基础模式的近似部分一起作为模型的输入来求解下一时刻的最优解，同时还利用历史参考点对预测值进行修正，从而提高了预测的准确性。

2　OiKF 短期交通预测理论模型及实现

在理论模型中，首先利用离散小波变换对异常值进行过滤处理，然后分析卡尔曼滤波器(KF)进行交通流预测的通用方法，最后提出基于异常值识别卡尔曼滤波器(OiKF)及其理论模型，并叙述利用该理论模型实现短时交通流预测的过程。

2.1　离散小波变换对异常值的处理

小波变换(wavelet transform，WT)是通过伸缩和平移对信号进行细化分析的工具，能提供变化的“时间-频率”窗口[16]。设函数 $\psi(t)$ 是一个平方可积的函数，即 $\psi(t)\in L^2(R)$，且该函数的傅里叶变换为

$$\Psi(\omega)=\int_{-\infty}^{\infty}\psi(t)e^{-j\omega t}dt$$

满足容许条件(admissibility condition)即：

$$C_\Psi=\int_{-\infty}^{\infty}\frac{|\Psi(\omega)|^2}{|\omega|}d\omega<+\infty$$

则称函数 $\psi(t)$ 为小波母函数，对其进行伸缩和平移变换得

$$\psi_{a,b}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), a>0, b\in R$$

其中 a 为尺度参数，b 为平移参数，若信号 $f(t)$ 是一个平方可积函数，即 $f(t)\in L^2(R)$，则连续小波变换的定义如下：

$$W_f^{\psi}(a,b)=\int_{-\infty}^{\infty}f(t)\bar{\psi}_{a,b}(t)dt, a>0, b\in R$$

其中，$\bar{\psi}_{a,b}(t)$ 为 $\psi_{a,b}(t)$ 的共轭。对尺度参数 a 和平移参数 b 进行离散化采样，并使两个参数满足 $a=a_0{}^j(a_0>1), b=ka_0{}^jb_0$，可获得离散小波函数：

$$\psi_{j,k}(t)=a_0{}^{-j/2}\psi(a_0{}^{-j}t-kb_0)$$

再令 $b_0=1$，使 $\psi_{j,k}(t)=a_0{}^{-j/2}\psi(a_0{}^{-j}t-k), a_0>1, j,k\in Z$，若信号 $f(t)\in L^2(R)$，则 $f(t)$ 的离散小波变换为

$$W_f^{\psi}(j,k)=\int_{-\infty}^{\infty}f(t)a_0{}^{-j/2}\bar{\psi}(a_0{}^{-j}t-k)dt$$

通常，$\psi_{j,k}(t)$ 是一个 $L^2(R)$ 上的标准正交基，即

$$\psi_{j,k},\psi_{m,n}=\begin{cases}1, j=m\text{且}k=n\\ 0, \quad \text{其他}\end{cases}$$

则，信号 $f(t)$ 可通过以下方式重构：

$$f(t)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}W_f^{\psi}(j,k)\psi_{j,k}(t)$$

另外针对离散时间信号的离散小波变换可通过若干次高通和低通滤波简单实现，其中高通滤波器获得高频信息，表示信号的细节，低通滤波器获得低频信息，表示信号的近似。低频信息和高频信息分别通过对应的合成滤波器，可以获得原始信号对应的低频部分和高频部分，再根据实际情况合成所需的部分，就可实现信号的分解和重构，如信号去噪、过滤异常值等。

2.2　KF 短期交通预测模型

已知 $H_t=[vol_{t-n},vol_{t-n+1},\ldots,vol_{t-1}]$：时刻 *t-n* 到时刻 *t-1* 的一组交通流原始数据，在短期交通预测中，预测目标值为 vol_t：即预测下一时刻 t 的交通流量。在 Zhou 等人[7]先前的工作中，将要预测的交通流量表示为过去交通流量和相应权重 $\Theta_t=[\theta_{t-n},\theta_{t-n+1},\ldots,\theta_{t-1}]^T$ 的组合，即：

$$vol_t=H_t\Theta_t+v_t \tag{1}$$

$$\Theta_t=F\Theta_{t-1}+w_t \tag{2}$$

其中 $F\in R^{n\times n}$ 表示状态转移矩阵，$v_t\in R^m$ 和 $w_t\in R^n$ 为测量噪声和过程噪声，它们均为零均值、不相关的高斯白噪声，分别有已知的过程噪声协方差矩阵 Q 和测量噪声协方差矩阵 R。一旦获得 vol_t，所求下一时刻 *t+1* 的交通流预测值 vol_{t+1} 可表示为

$$vol_{t+1}=H_{t+1}\Theta_t \tag{3}$$

Θ_t 是结合式(1)和(2)得到的时刻 t 最优估计值。Lan 等人[17]、Xie 等人[14]和 Zhou 等人[7]的工作也做了相同的假设。

2.3　OiKF 短期交通预测模型

OiKF 短期交通预测模型中，利用离散小波变换对交通流原始数据 $vol_t\in H_t$ 进行分解重构，最终由两部分组成：近似部分 $\tilde{H}_t$（$\tilde{H}_t=[vol_{t-n},vol_{t-n+1},\cdots,vol_{t-1}]$）和细节部分 Λ_t（$\Lambda_t=[\lambda_{t-n},\lambda_{t-n+1},\cdots,\lambda_{t-1}]$）。其基本原理为：由于异常值(存在高度不确定性：气候、交通事故、突发事件等因素在原始数据流中产生的异常值)的存在，原始交通流数据被局部存在的异常值/噪声严重破坏，为解决该问题，通过离散小波变换，以分解重构获得的低频近似部分 vol_t 替代原始数据 vol_t 作为系统的部分输入用于交通流预测，实现在对原始数据过滤由不确

定性产生的异常值，同时还保持了原始交通流的基础模式；为不丢失原始数据中指示交通流突变的细微线索，以另一部分——高频细节部分 Λ_t 作为被异常值所破坏的控制输入变量，去捕获隐含在快速随机变化的交通情况中的细微线索。

在离散小波变换对交通流原始数据 $vol_t \in H_t$ 进行分解重构的基础上，实现扩展卡尔曼滤波器，其动态线性系统可表示为

$$\tilde{\Theta}_t = F\tilde{\Theta}_{t-1} + B\Lambda_{t-1} + w_t \quad (4)$$

$$vol'_t = \tilde{H}_t\tilde{\Theta}_t + v_t \quad (5)$$

其中 $B \in R^{n \times n}$ 是控制矩阵，其作用与 F 类似。为了区别于式(2)中交通流权重 Θ_t，将其重新表示为 $\tilde{\Theta}_t$。细节部分 Λ_t 可被重新表示为

$$\Lambda_t = \Lambda_{t-1} + \xi_t \quad (6)$$

其中 $\xi_t \in R^n$ 表示协方差为 σ^2 的高斯白噪声。通过扩展，该技术既继承了传统卡尔曼滤波过滤噪声的所有优点，又保证了原始数据的质量。

在预测模型中，其目标成本函数为

$$\mathrm{E}(vol_t) = Cov(e_{t-n}^2 + e_{t-n+1}^2 + \ldots + e_{t-1}^2) - \sigma^2 K_t^2 \quad (7)$$

问题求解：利用公式(7)所示的成本函数去求解最优解。其中：$\mathrm{E}(vol_t)$ 表示时刻 t 求解获得的优化目标，Cov 运算表示每个时刻 i 的估计误差的平方 e_i^2 组成的协方差矩阵，$K_t \in R^{n \times n}$ 是卡尔曼增益矩阵。由于篇幅问题，推导优化公式(7)的详细过程可参考 Zhang 等人[8], Ma 等人[18]的相关工作。本文给出了该算法的基本过程：

首先基于 t-1 时刻的权重 $\tilde{\Theta}_{t-1}^+$ 和误差协方差矩阵 P_{t-1}^+ 得到一个先验的估计值 $\tilde{\Theta}_t^-$ 以及相应的误差协方差矩阵 P_t^-。然后用求得的卡尔曼增益矩阵 K_t 和测量值 vol_t 去更新估计值 $\tilde{\Theta}_{t-1}^+$ 和相应协方差矩阵，从而获得其后验估计 $\tilde{\Theta}_t^+$(所求的最优估计值)和相应协方差矩阵 P_t^+，在得到最优估计值后，所求的交通流预测值 vol_{t+1} 可通过公式(3)进行计算。最后 $\tilde{\Theta}_t^+$ 和 P_t^+ 下一时刻的输入(先验估计)用来计算下一次的最优估计，这是一个动态滚动的持续预测过程。

此外，通过对实际数据的波形图进行分析，可知交通流数据本质上具有一定的周期性。因此，历史数据具有很好的参考价值。在进行预测时，首先通过公式(3)获取时间 t 的预测值 vol'_t，然后通过历史参考值对预测值进行修正。在近似部分中，先选取时间 t 之前一段时间共 s 个时间点的数据作为当前序列 $S_t = [vol_{t-s}, vol_{t-s+1}, \ldots, vol_{t-1}]$，再从已预测获得的历史数据序列中搜寻与当前实际交通流数据序列最为接近的历史序列 $S_h = [vol_{h-s}, vol_{h-s+1}, \ldots, vol_{h-1}]$，选取时间 h 观测值 vol_h 作为参考值，预测值修正过程如图 1 所示。

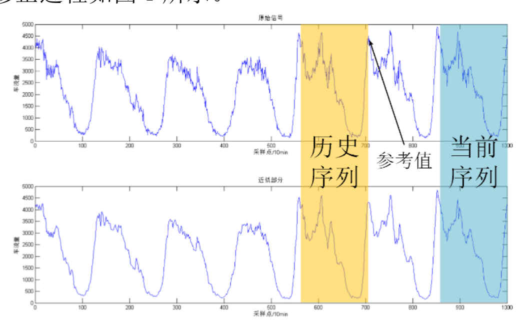


图 1 卡尔曼滤波器的预测值进行修正处理过程

Fig. 1 The correction process of prediction value of Kalman filter

vol_h 与 vol'_t 进行加权平均，得到最终的预测值 vol_t，设参考值的权重为 $weight$，其计算公式可定义为

$$vol_t = weight \times vol_h + (1 - weight) \times vol'_t \quad (8)$$

基于异常值识别卡尔曼滤波器短期交通预测基本过程为

算法 1 基于异常值识别卡尔曼滤波器算法步骤

1．初始化阶段：

1.1 设置状态转移矩阵 F，控制矩阵 B，过程噪声的协方差矩阵 Q，测量噪声的协方差矩阵 R 和细节部分的噪声方差 σ^2。

1.2 初始化 $\tilde{H}_{t_0}$ 状态协方差矩阵 P_{t_0}，权重 $\tilde{\Theta}_{t_0}$ 和细节部分 Λ_{t_0} 的值。

2．预测阶段：

$\tilde{\Theta}_t^- = F\tilde{\Theta}_t^+ + B\Lambda_{t-1}$

$P_t^- = FP_{t-1}^+F^T + \sigma^2 BB^T + Q$

3．更新阶段：

$K_t = P_t^-\tilde{H}_t^T\left(\tilde{H}_t P_t^- \tilde{H}_t^T + R - \sigma^2\right)^{-1}$

$\tilde{\Theta}_t^+ = \tilde{\Theta}_t^- + K_t\left(vol_t - \tilde{H}_t\tilde{\Theta}_t^-\right)$

$P_t^+ = \left(I - K_t\tilde{H}_t\right)P_t^-$

$vol'_t = \tilde{H}_t\tilde{\Theta}_t + R$

4．修正阶段：

搜寻最相似的历史序列，获得历史参考值 vol_h，根据公式(8)与第 3 阶段获得的预测值 vol'_t 进行加权平均，得到最终的预测值 vol_t。

5．迭代：重复阶段 2~4 直到预测完毕。

3 实验及结果对比分析

3.1 实验数据说明

实验中采用了常被用作基准的测试数据——荷兰阿姆斯特丹的四条高速公路的交通流量数据(分别表示为 A1、A2、A4 和 A8)[4,5,7,8,15,19]，对本文提出的 OiKF 进行验证性实验。这四条高速公路终止于阿姆斯特丹的环路，即 A10 高速公路，对应的分布如图 2 所示，其数据是通过分别位于距离环路融合点附近的一段距离内的 MONICA 传感器的四个检测点收集获得。数据序列为：从 2010 年 5 月 20 日至 2010 年 6 月 24 日，每分钟进行一次统计，统计该分钟内所经过的车辆数量，并折算成该分钟的每小时车流量。

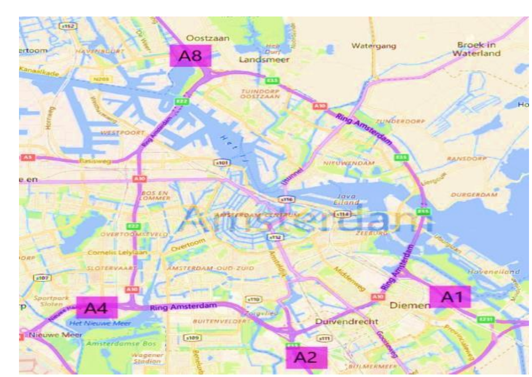


图 2 荷兰阿姆斯特丹的 A1、A2 等四条高速公路分布图

Fig. 2 The map of the four highways in Amsterdam

这四条高速路具有一定的代表性，其中：

a) A1 为边境高速，是欧洲第一条高占有率 3+分离车道，车道占有率瞬时变化较大，预测难度高。

b) A2 为荷兰境内最繁忙的高速公路之一，其拥堵情况变化大，用以验证预测模型应对交通拥塞的预测有效性。

c) A4 高速公路是境内标准的国家公路，其车流情况比较缓和正常。

d) A8 高速公路是一段长约 10 公里左右的接驳性公路，以验证交通突发情况的预测有效性。

3.2 实验设置说明

在实验中，令状态转移矩阵 F 的初始值设置为单位矩阵 I，即 $F = I$；控制矩阵 B 初始设置为 $10^{-5} \cdot I$ 即 $B = 10^{-5} \cdot I$；矩阵大小 n 设置为 8。测量异常值的协方差矩阵 R 设置为零矩阵，初始状态协方差矩阵 P_t 和过程异常值的协方差矩阵 Q 设置为 $10^{-2} \cdot I$。令初始权重 Θ_t 设置为 $\left[\frac{1}{n}, \cdots, \frac{1}{n}\right]$。另外在离散小波变换操作时，本文选择 Daubechies 1 和 Haar 作为母波。在搜索历史参考值时，经过实验验证，按一天时间

($24\times60min=144\times10min$)进行切分的数据序列相关性最高，因此将序列的长度 s 设置为 144，历史参考点的权重 $weight$ 的值设置为 0.2。

3.3　实验评价准则

在实验中，拟定预测效果评估准则为[4, 5,7, 8,14,19~23]：均方根误差(root mean square error，RMSE)用于衡量一个模型的预测值和系统的测量值之间的平均误差；平均绝对百分比误差(mean absolute percentage error，MAPE)是这个平均误差的百分比表示。其定义为

$$\mathrm{RMSE}=\sqrt{\frac{1}{M}\sum_{t=1}^{M}\left(vol_t-vol_t\right)^2} \tag{9}$$

$$\mathrm{MAPE}=\frac{1}{M}\sum_{t=1}^{M}\left|\frac{vol_t-vol_t}{vol_t}\right|\times100\% \tag{10}$$

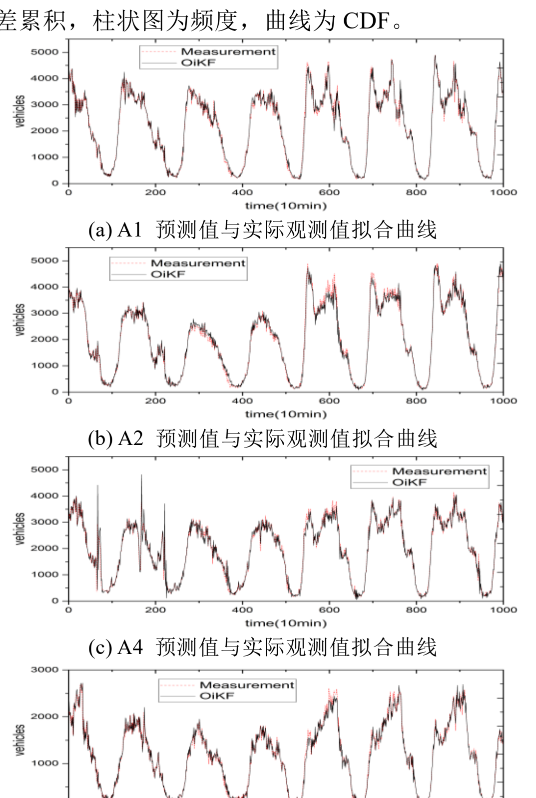
其中 M 表示数据集的大小，vol_t 和 vol_t 分别为时刻 t 的预测值和真实值。

3.4　预测结果及分析

实验采用 OiKF 对数据集进行测试。图 3(a)-(d)为针对 A1，A2，A4，A8 四条公路交通流量数据其中 1000 个点的预测结果。图中黑色实线为预测值，红色虚线为实际观测值。图 4-7 为对应 A1，A2，A4，A8 四条公路交通流量 5000 个点的预测值与实际观测值的相关误差频度及 CDF 图，由观测值与预测值的差值取绝对值并除以观测值获得。

实验数据集考虑了多种影响因素及环境下的交通流数据，如图 3 所示，预测曲线与实际观测值曲线的拟合度高，即使在数据突变的位置误差也比较小。由此可见本文提出的方法克服了传统卡尔曼滤波器的不足，有效地捕捉到指示交通流突变的细微线索，能准确地对短时交通流进行预测。

图 4~7 中，x 轴表示误差区间，左边 y 轴表示 5000 个预测点中误差 $\left|\frac{vol_t-vol_t}{vol_t}\right|$ 落入对应区间的数量(频度)，右边 y 轴表示误差累积，柱状图为频度，曲线为 CDF。



(a) A1　预测值与实际观测值拟合曲线

(b) A2　预测值与实际观测值拟合曲线

(c) A4　预测值与实际观测值拟合曲线

(d) A8　预测值与实际观测值拟合曲线

图 3　四条高速公路交通流量预测值与实际观测值拟合曲线图

Fig. 3　The curve of the predictions and actual values

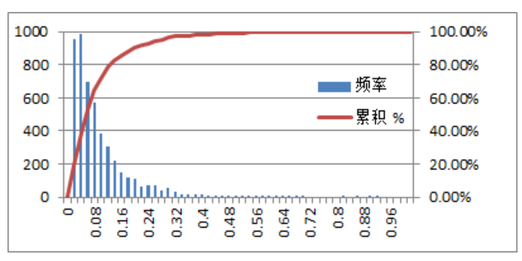


图 4　A1 预测误差值频度及误差 CDF

Fig. 4　The prediction error frequency and error CDF of A1

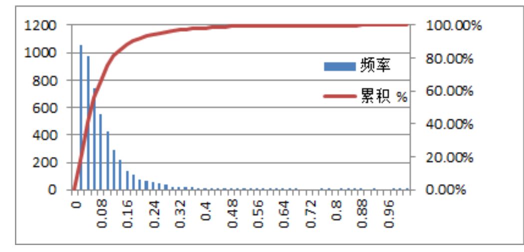


图 5　A2 预测误差值频度及误差 CDF

Fig. 5　The prediction error frequency and error CDF of A2

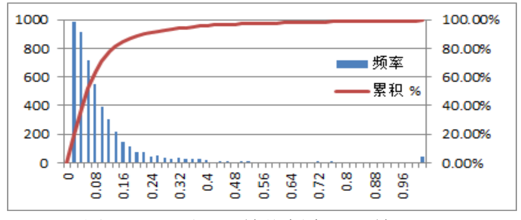


图 6　A4 预测误差值频度及误差 CDF

Fig. 6　The prediction error frequency and error CDF of A4

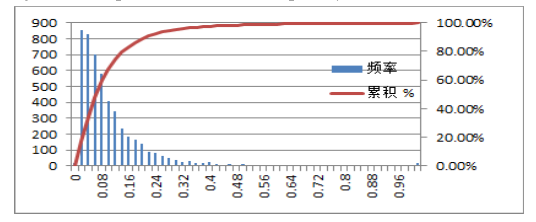


图 7　A8 预测误差值频度及误差 CDF

Fig. 7　The prediction error frequency and error CDF of A8

通过分析 5000 个数据点预测误差落入的区间频度和累积 CDF 曲线，a)由图 4 得到：对 A1 预测，误差在[0,0.2]累积达到 90.36%，误差<0.3 累积达到 96.39%；b)对于 A2，由图 5 可知：预测误差在[0,0.2]累积达到 92.03%，误差<0.3 累积达到 96.55%；c)图 6 可知，对于 A4 的交通预测，[0,0.2]累积达到 88.56%，误差<0.4 累积达到 95.95%；d)图 7 所示，[0,0.2]累积达到 89.10%，误差<0.4 累积达到 95.45%。所有的预测中，误差<=0.2 都超过了 90%，证明了 OiKF 能准确地对短时交通流进行预测。

3.5　预测准确性对比分析

为验证 OiKF 模型在整体预测的有效性，以同一数据集(荷兰阿姆斯特丹的四条高速公路的交通流量数据)为输入，将本文提出的 OiKF 预测结果数据与传统的卡尔曼滤波模型[11] 以及当前一些新提出的交通流预测模型的实验结果按评价准则——式(8)(9)的 MAPE/RMSE 进行准确性对比分析。主要具有代表性的对比算法有：

a) 混合粒子群优化支持向量机回归模型算法(hybrid particle swarm optimization support vector regression method，SVR)[11]；

b) 灰色模型算法(grey model，GM)[12]；

c) 常用的标准基线对照模型算法[24]：随机游走模型(random walk method，RW)、历史平均模型(historical average model，HA)、基于机器学习的最小平方提升模型(least squares boosting，LSBOOST)，一个简单模型的集成框架[13]和指数平

滑模型(exponential smoothing model，ES)；

d) 当前一些新的研究成果的模型算法：自动回归模型算法(autoregression，AR)[4]、混合式双卡尔曼滤波模型(hybrid dual Kalman filtering model，KF)[25]、基于混合学习的模型(hybrid learning based model，SVRGSA 以及 SVRPSO)[19]。

各类算法的 MAPE 和 RMSE 实验结果数据分别如表 1 和 2 所示，最后本文还分析了 OiKF 对比各类算法在有效性上的提升率。

表 1　本方法和当前各类预测模型的 MAPE 实验结果对比

Tab. 1　The comparison of MAPE with other models

model	A1	A2	A4	A8
SVR	14.34	12.22	12.23	12.48
GM	12.55	10.88	13.28	12.92
RW	12.65	11.43	12.06	12.37
HA	16.87	15.53	16.72	16.24
LSBOOST	13.78	14.43	12.90	14.00
ES	11.94	10.75	11.97	12.00
AR	13.57	11.59	12.70	12.71
KF	12.46	10.72	12.62	12.63
SVRGSA	11.15	9.42	10.65	11.81
SVRPSO	11.63	10.08	10.99	12.20
OiKF	**8.57**	**7.93**	**10.61**	**11.56**

表 2　本方法和常用预测方法的 RMSE 实验结果对比

Tab. 2　The comparison of RMSE with other models

model	A1	A2	A4	A8
SVR	329.09	259.74	253.66	190.30
GM	348.38	255.86	274.48	188.97
RW	312.92	223.82	230.01	174.14
HA	404.84	348.96	357.85	218.72
LSBOOST	306.33	233.88	229.78	177.52
ES	315.82	226.40	237.76	174.67
AR	301.44	214.22	226.12	166.71
KF	332.03	239.87	250.51	187.48
SVRGSA	284.97	192.68	213.69	161.07
SVRPSO	300.97	205.94	224.63	163.95
OiKF	**203.34**	**154.98**	**184.96**	**132.41**

从表 1 和 2 可以得出，本文提出的方法在四个基准数据集上，比其他的对照方法更准确，OiKF 在 MAPE 和 RMSE 的实验结果中都取得了最佳效果，与其他各个模型算法预测结果在 MAPE 和 RMSE 上的效果提升情况如图 8 和 9 所示。

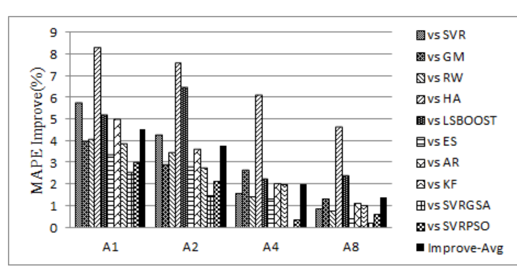


图 8　与其他模型在 MAPE 上的误差降低对比

Fig. 8　The comparison of MAPE improving with others model

由图 8 可知，OiKF 模型在 A1、A2、A4 和 A8 四条高速公路的测量点的预测值所获得的 MAPE 是最好的。图 8 中 Impvoe-Avg 项统计出 OiKF 模型在各条高速路上对比其他预测模型的 MAPE 的平均降低百分率，故由图 8 可知，本文所提出的预测模型比其他最新研究提出的预测模型(如 AR，SVRGSA，SVRPSO 等)，其误差在四条高速路的预测值误差平均分别降低了达到：4.524%，3.775%，2.002%和 1.376%。

分析图 9 可知，与常用及当前最新研究的预测模型比较，在 RMSE 的 4*12 组的预测结果数据对比分析得到，OiKF 所获得的效果也是最好的，从图 9 中 Impove-Avg 项数据可知，OiKF 的 RMSE 值整体上比其他模型相应地平均分别降低了 120.339，85.157，64.889 和 47.943。

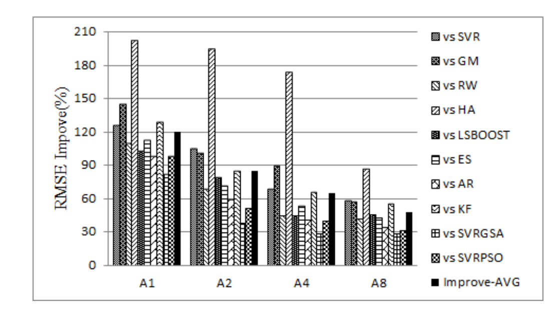


图 9　与其他模型 RMSE 值降低值对比

Fig. 9　The comparison of RMSE improving with others model

OiKF 在四条高速路上预测结果的 MAPE 和 RMSE 整体提升情况(与其他模型对比，MAPE 降低值与 RMSE 减少值的分布图)对应的箱图如图 10 和 11 所示。

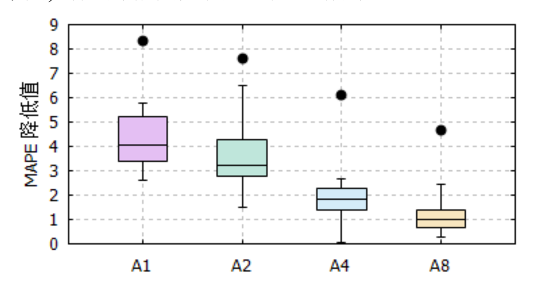


图 10 各路线预测 MAPE 降低值分布情况

Fig. 10　The distribution of MAPE improving values

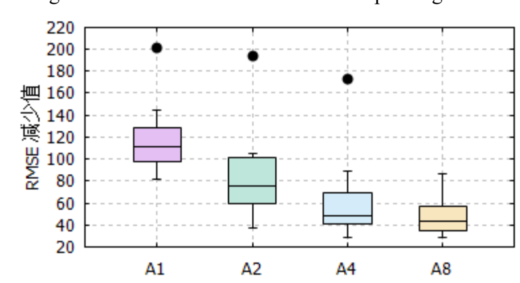


图 11 各路线预测 RMSE 减少值分布情况

Fig. 11　The distribution of RMSE improving values

通过对实验结果的比较分析，实验数据呈现出本文提出的预测模型比其他常用模型的优越性，也充分证明了本文提出的方法在解决短期交通流预测问题中的有效性。

4　模型应用实例

在解决公交企业目前普遍存在的“分散调度”、“静态调度”、“经验调度”，引起的投入高、反应慢、效益差、客源流失严重的问题时，将基于异常值识别卡尔曼滤波器的短期交通流预测模型嵌入在公交智能调度系统中，对特殊路段进行短期交通流量预测，以动态调整公交调度的需求，提高公交运营调度管理的智能化，实现公交企业对人、车、线路等资源的高效利用。图 12 为系统中公交调度与 GIS 整合的场景，其中利用 OiKF 对地图中的多条主要路径进行交通流量预测，并反馈到公交的实时调度中。

本文提出的 OiKF 模型测试应用在 2019 年工业和信息化部新型信息消费示范项目“桂林出行网——运输服务与旅游融合数字经济服务平台”的公交智能调度系统中，分别在桂林市的环城北一路、环城北二路、中山北路、东二环路和建干路五条干道(如图 12 所示标注的五条城市干道)上设置了 1000 个预测点，以对 OiKF 进行有效性验证测试，并获得了

良好的运营效果。将前期预测情况反馈给调度系统，以预案有效地指挥、控制和调节车辆运行，均匀合理的行车间隔，营运车辆跨线路营运，并可用于回顾总结分析，实现调度业务管理闭环。

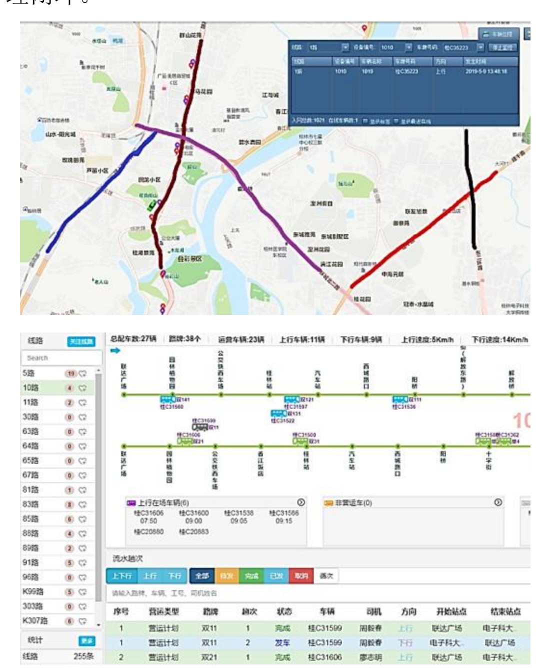


图 12　OiKF 应用实例场景图

Fig. 12　Oikf in the bus intelligent dispatching system

5　结束语

本文提出了一种新的短时交通流预测方法，用于改善短期交通流预测的准确性。该方法改善了传统的卡尔曼滤波器，使其能对异常值进行识别和过滤。通过实验，与当前研究成果提出的模型的对比以及在公交智能调度系统中的测试应用，实验结果证明了该方法有效性和优越性，也说明在进行短期交通流预测时，过滤异常值的同时保证原始数据的质量是必要的。在四个基准数据集上的大量实验表明，与常用及最新的预测模型相比，其结果 MAPE 平均降低了 2.919%，RMSE 平均降低了 79.582。在今后的研究中，将进一步研究扩展到其他卡尔曼滤波器中，如扩展卡尔曼滤波器或无迹卡尔曼滤波器，并将其应用于其他预测任务中，如服务器任务序列的分析和预测，云计算中心各类资源利用率预测和分析，以及小粒度任务调度或任务量随时间变化预测等。结合神经网络的应用和一些新的交通流短期预测模型的成果[26~28]，将进一步优化 OiKF 模型的参数设置，以提高预测模型的准确性。

参考文献：

[1] Teresa P. Impact of Data Loss for Prediction of Traffic Flow on an Urban Road Using Neural Networks [J]. IEEE Transactions on Intelligent Transportation Systems, 2018: 1-10.

[2] Zhan Hongyuan, Gomes G, Li X S, *et al.* Consensus ensemble system for traffic flow prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19 (12): 3903-3914.

[3] Zheng Zibin, Yang Yatao, Liu Jiahao, *et al.* Deep and embedded learning approach for traffic flow prediction in urban informatics [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20 (10): 3927-3939.

[4] Zhou Teng, Han Guoqiang, Xu Xuemiao, *et al.* δ-agree AdaBoost stacked autoencoder for short-term traffic flow forecasting [J]. Neurocomputing, 2017, 247: 31-38.

[5] Zhou Teng, Han Guoqiang, Xu Xuemiao, *et al.* A Learning-Based Multimodel Integrated Framework for Dynamic Traffic Flow Forecasting [J]. Neural Processing Letters, 2018 (5): 1-24.

[6] Wang Yubin, Van Schuppen J H, Vrancken J. Prediction of traffic flow at the boundary of a motorway network [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 15 (1): 214-227.

[7] Zhou Teng, Jiang Dazhi, Lin Zhizhe, *et al.* Hybrid dual Kalman filtering model for short-term traffic flow forecasting [J]. IET Intelligent Transport Systems, 2019, 13 (6): 1023-1032.

[8] Cai Lingru, Zhang Zhanchang, Yang Junjie, *et al.* A noise-immune Kalman filter for short-term traffic flow forecasting [J]. Physica A: Statistical Mechanics and its Applications, 2019, 536: 122601.

[9] Okutani I, Stephanedes Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research Part B: Methodological, 1984, 18 (1): 1-11.

[10] Stathopoulos A, Karlaftis M G. A multivariate state space approach for urban traffic flow modeling and prediction [J]. Transportation Research Part C: Emerging Technologies, 2003, 11 (2): 121-135.

[11] Hu Wenbin, Yan Liping, Liu Kaizeng, *et al.* A short-term traffic flow forecasting method based on the hybrid PSO-SVR [J]. Neural Processing Letters, 2016, 43 (1): 155-172.

[12] Guo Huan, Xiao Xinping, Jeffrey F. Urban road short-term traffic flow forecasting based on the delay and nonlinear grey model [J]. Journal of Transportation Systems Engineering and Information Technology, 2013, 13 (6): 60-66.

[13] Mayrink V, Hippert H S. A hybrid method using Exponential Smoothing and Gradient Boosting for electrical short-term load forecasting [C]// 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI) . IEEE, 2016: 1-6.

[14] Xie Yuanchang, Zhang Yunlong, Ye Zhirui. Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition [J]. Computer-Aided Civil and Infrastructure Engineering, 2007, 22 (5): 326-334.

[15] Wang Yubin, Van Schuppen J H, Vrancken J. Prediction of Traffic Flow at the Boundary of a Motorway Network [J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15 (1): 214-227.

[16] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis [J]. Journal of Renewable & Sustainable Energy, 1990, 36 (5): 961-1005.

[17] Lin Weihua. A Gaussian maximum likelihood formulation for short-term forecasting of traffic flow [C]// Intelligent Transportation Systems. IEEE, 2001: 150-155.

[18] Ma Wentao, Qiu Jinzhe, Liang Junli, *et al.* Linear Kalman Filtering Algorithm with Noisy Control Input Variable [J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2018: 1-1.

[19] Cai Lingru, Chen Qian, Cai Weihong, *et al.* SVRGSA: a hybrid learning based model for short-term traffic flow forecasting [J]. IET Intelligent Transport Systems, 2019, 13 (9): 1348-1355.

[20] Mackenzie J, Roddick J F, Zito R. An evaluation of HTM and LSTM for short-term arterial traffic flow prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20 (5): 1847-1857.

[21] Yang Haofan, Dillon T S, Chang E, *et al.* Optimized configuration of exponential smoothing and extreme learning machine for traffic flow forecasting [J]. IEEE Transactions on Industrial Informatics, 2018, 15 (1): 23-34.

[22] Feng Xinxin, Ling Xianyao, Zheng Haifeng, *et al.* Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20 (6): 2001-2013.

[23] Zheng Chuanpan, Fan Xiaoliang, Wen Chenglu, *et al.* Deepstd: Mining spatio-temporal disturbances of multiple context factors for citywide traffic flow prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2019.

[24] Lippi M, Bertini M, Frasconi P. Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14 (2): 871-882.

[25] Zhou Teng, Jiang Dazhi, Lin Zhizhe, *et al.* Hybrid dual Kalman filtering model for short-term traffic flow forecasting [J]. IET Intelligent Transport Systems, 2019, 13 (6): 1023-1032.

[26] 王秋莉，李军. 基于核学习方法的短时交通流量预测 [J]. 计算机应用研究, 2019 (3): 696-700. (Wang Qiuli, Li Jun. Short-term traffic flow forecasting based on kernel learning mathods [J]. Application Rsearch of Computers, 2019 (3): 696-700.)

[27] 曹堉，王成，王鑫，高悦尔. 基于时空节点选择和深度学习的城市道路短时交通流预测 [J/OL]. 计算机应用: 1-10. http://kns. cnki. net/kcms/detail/51. 1307. TP. 20191209. 1313. 008. html. (Cao Yu, Wang Cheng, Wang Xin, *et al.* Urban road short-term traffic flow prediction based on spatio-temporal node selection and deep learning [J/OL]. Journal of Computer Applications, pp 1-10. http://kns. cnki. net/kcms/detail/51. 1307. TP. 20191209. 1313. 008. html.)

[28] 林浩，李雷孝，王慧. 支持向量机在智能交通系统中的研究应用综述 [J/OL]. 计算机科学与探索: 1-19. http://kns. cnki. net/kcms/detail/11. 5602. tp. 20200331. 1857. 004. html. (Lin Hao, Li Leixiao, Wang Hui. A Survey on research and application of support vector machines in ITS [J/OL]. Journal of Frontiers of Computer Science and Technology, pp 1-19. http://kns. cnki. net/kcms/detail/11. 5602. tp. 20200331. 1857. 004. ht ml.)